White Paper

Intel Information Technology

Computer Manufacturing

Datacenter Management

Building an Enterprise Data Warehouse and Business Intelligence Solution

To achieve end-to-end visibility into critical business functions across the company, Intel IT deployed an integrated warehouse solution. The solution employs a consolidated enterprise data warehouse (EDW) along with business intelligence (BI) applications. The project was a multi-team, cross-business-unit effort that began in 1999 and went online in 2001, migrating over 200 individual datamarts spread across multiple business units around the globe. Today's EDW occupies over 20TB of usable data space, processes 90 billion rows of data per month, and is accessed by over 11,000 users worldwide—resulting in a cumulative business benefit in the hundreds of millions of dollars.

Charles Eden and Venky Padmanabhan, Intel Corporation

February 2006

IT@Intel

Executive Summary

Intel IT partnered with multiple Intel business units around the globe to migrate over 50 applications and 200 individual datamarts into a consolidated enterprise data warehouse (EDW) that the entire company uses today. The data warehouse currently occupies over 20TB of data space, processes over 90 billion rows of data per month, is accessed by over 11,000 users worldwide, and has delivered hundreds of millions of dollars in value to Intel's bottom line.

More effective operations allowed us to achieve a cumulative business value in the hundreds of millions of dollars from 2001 through 2005. We were able to reduce and avoid IT-related operating costs of over $22 million in 2005 alone.

Critical to the success of delivering a solution was early engagement with key stakeholders and involving them in vendor selection, discovery, migration planning, and implementation.

The solution offers:

- **Scalability.** Since deployment, we've expanded from just over 100 users running 30 applications on 317 databases, to over 11,000 users, 284 applications, and over 2,500 databases.
- **Cost savings.** More effective operations allowed us to achieve a cumulative business value in the hundreds of millions of dollars from 2001 through 2005. We were able to reduce and avoid IT-related operating costs of over $22 million in 2005 alone.
- **Analytics/data mining.** In-depth, multi-dimensional analytics deliver end-to-end visibility and manageability of products and processes across the company.
- **Reporting tools.** Standardized analysis and reporting tools enable quick and easy access to business-critical data worldwide and facilitate Intel compliance with the Sarbanes-Oxley Act of 2002 (SOX). Automated reporting for SOX saves millions.
- **Security.** Advanced security capabilities and technology fortification help provide data integrity, security, and protection.
- **Performance.** The high-performance, Intel® Xeon® processor-based, massively parallel processing (MPP) architecture speeds time-to-query results from hours on the old systems to minutes or seconds today. While our old system could sometimes be down for days, uptime with the EDW solution is 99.99 percent, allowing us to maintain throughput and exceed customer expectations.

Contents

Business Challenge

Intel business units access terabytes of data every day to make and execute critical business decisions and track activities that affect spending and profits. From daily operations to quarterly financial reports, the data we gather and manage in our data warehouse is critical to running an expanding enterprise. These business units depend on the quality and integrity of business data, and their ability to effectively access and derive critical information from it.

An Evolving Problem

In 1993, Intel IT began hosting databases for the company's business units, which grew to over 200 expanding datamarts by 1998. We were responsible for providing business intelligence (BI) services—extracting data and creating analysis applications that output reports for business unit consumption. Each business unit maintained one or more databases, often with different data models. In the late 1990s, the complexity of this infrastructure began to manifest the following issues:

- **Productivity.** Business unit users spent increasing amounts of time getting tasks done, such as financial year-end closings.
- **Verification.** Users and BI analysts had to cross-check multiple data sets to verify integrity across so many data sources.
- **Costs.** Managing a vast repository infrastructure and duplicate data was costly.
- **Extract, transform, and load.** The time it took to extract, transform, and load (ET&L) data from disparate data warehouses was unacceptable. For example, analysts typically spent 30 percent of their time just gathering information from multiple sources, verifying its integrity, and transforming it to meet their needs—before actually doing any analysis.

Bigger Concerns

Managers across the company were concerned not only about productivity and costs, but also about the potential loss of business opportunities caused by the lack of end-to-end visibility and multi-dimensional analytics.

From an IT operations perspective, the infrastructure was becoming costly. As business units added new applications, costs not only grew, they multiplied, because of many instances of duplicate applications and data. This led to quality concerns as we compared one set of data to another.

We also needed to optimize the use of our mixed data warehouses, such as for end-to-end supply-chain management and detailed sales and marketing analytics, to run the business more effectively.

The technology we were using in 1998 didn't allow us to simply pick one of our existing systems and then convert and scale it to meet all our needs. By this time, EDW and BI solutions had become powerful solutions and were receiving the attention of large businesses and analysts. We determined that we needed to launch a unified, scalable EDW with BI capabilities that would meet the needs of our many business units around the globe.

Our Opportunity

A unified EDW would make IT operations more efficient. With a single infrastructure and data set to manage, we would reduce the costs associated with data duplication and redundant systems. Consolidation would also allow us to reuse existing data for new applications, which could reduce possible data quality problems.

From a business perspective, the value of end-to-end visibility and manageability from the multi-dimensional analysis that BI offers was considerable. Business units increasingly sought information about how activities in other business units affected their own. In addition, not having to convert data from other repositories meant faster answers and more efficient business activity.

With an EDW and BI applications, we had a clear opportunity to significantly reduce current and future operating costs, avoid data quality complexities, and bring value to the company's bottom line.

Determining the Solution

We collaborated with business units and went through a phased discovery process to define our requirements, select a vendor, and deploy the system.

Discovery and Definition

We formed a cross-functional team drawn from the IT organization and the business units that would depend on the new data warehouse. By surveying business unit managers and key stakeholders about the issues and concerns they faced with the existing data warehouses, we received feedback that defined and directed requirements. Feedback covered various areas:

- **Data quality.** Business units each had their own method of defining what the data should look like (its metadata), which would require different manipulation for ET&L, possibly leading to multiple results in the data that was finally loaded. Duplication of data across datamarts needed to be validated before it could be trusted. Impending new regulatory requirements also emphasized the importance of data quality, management, and retention.
- **Performance and productivity.** Due to inadequacies in the aging legacy technology, the time it took for ET&L and report queries was unacceptable. These issues acutely affected business units with strict deadlines for posting their results.
- **Cost.** Multiple datamarts across the company resulted in duplicate infrastructure and staff, making costs a serious consideration. Potential losses from lack of end-to-end visibility and management of data quality and our processes also posed concerns.

Defining IT Requirements

The team worked diligently over the course of a year to complete a requirements document satisfactory to key stakeholders. The following were among our infrastructure requirements:

- **Scalability.** We needed to be able to easily expand the EDW platform without impacting productivity, manageability, or data integrity.
- **Performance.** Because productivity was a top concern for users, high throughput was essential. In addition, we needed to be able to scale without compromising performance. We realized that these capabilities depended on system architecture, both hardware and software, so we needed a solution that would allow us to manage performance and not limit our ability to scale as Intel grew.
- **Data integrity and security.** Data is a crucial asset. Maintaining its integrity and security is a business-critical function of IT. We needed a flexible, cost-effective solution that protected data and the environment that hosted the EDW and BI applications.
- **Vendor capability.** This was not a turnkey project. Our data repositories were diverse, complex, and specialized, and we expected our needs would further evolve. We needed vendors with competencies that complemented our own expertise to work with us to provide the tools, high availability, and security we required as part of a scalable, high-performance solution.

Identifying Business Needs

In addition to infrastructure requirements, we also created a list of business capabilities. Important applications included the following:

- **Multi-dimensional analytics.** With the ability to analyze data across all geographies, product lines, cost departments, and so on, managers could make better informed and intelligent decisions about their business units' operations and optimize functions accordingly.
- **Data mining.** Over the years, the more data analysts extracted and analyzed, the more valuable the data became—especially data that could be mined across business units. However, extraction and transformation took increasingly longer because of the data quantity and quality, and the different data models used across the company. This limited our ability to make quick decisions about possible opportunities. A single EDW with effective mining capabilities could increase our agility.

Achieving Agreement

Once we fully characterized the solution, we needed to get agreement from all the business units that would be affected by the new system—in terms of validating their requirements, changing their processes, and finally, to secure the funding and commitment required to migrate their data. This was a critical goal for the success of a unified EDW.

Vendor Selection and Solution Benchmark

The selection and benchmarking process took approximately eight months.

Vendor Selection

We considered solutions from six major system providers, first by evaluating their product specifications. We also interviewed other companies and IT organizations about their experiences within the data warehouse industry. From published specifications and other documents, we quickly eliminated three of the possible vendors because they didn't fully meet our major criteria of scalability, performance, and vendor capabilities.

To validate the claims of the remaining three system providers, we assembled problem statements and reports we required from our existing datamarts, and then pilot tested them in the providers' labs. After conducting performance tests—including query, data loading, data transformation, and back up times; throughput; data validation; and re-indexing—and evaluating what was required to add enterprise-level capacity, one likely provider emerged.

Solution Benchmark

Using our existing datamarts, we benchmarked certain data sets and applications with which we'd already had experience and then compared against the new solution. However, benchmarking was not simply a matter of loading existing data into the development system and running queries. We were not only evaluating a new system, we were designing an entire data warehouse test bed using our existing data. We had to define the warehouse metadata, load the test data, and understand how we would run the queries on the new system.

Throughout our benchmarking, we continually stressed the system by adding new capabilities to see how they would impact data security, performance, data loading time, and query execution time. With each test, we measured and tracked data load time and query response time. Our final analysis also included acquisition and management costs.

Benchmark results revealed a system that had the capabilities we needed. Performance improvements were significant: queries that used to take hours ran in minutes—or seconds. The system's technology enabled it to meet our IT requirements and business needs.

Enterprise Data Warehouse and Business Intelligence Technology

We needed to ensure that the system would provide scalability, data integrity, security, and high availability. It needed to easily grow with Intel, so we made sure it was based on the latest and fastest technologies in system architecture, platform and processor performance, fiber optic interconnects, high-speed input and output, and disk subsystems. To help manage costs over time, we needed the solution to integrate with multiple generations of hardware—for example, from single-core to dual-core and multi-core Intel® processor-based platforms. We also wanted it to support a single relational database management system (RDBMS) software solution optimized on a standardized, scalable platform.

Enterprise Data Warehouse Architecture

The architecture incorporates multiple low-profile, rack-mounted servers running dual Intel® Xeon® processors linked via gigabit networks to create an MPP system with high-performance interconnects (see Figure 1). Redundancy is built in at several levels, including all network and disk subsystem interconnects, for high availability. The system is fully scalable to over 1,000 server nodes, and it is backwards compatible with three generations of hardware to help preserve existing capital investments and ease integration and support costs.

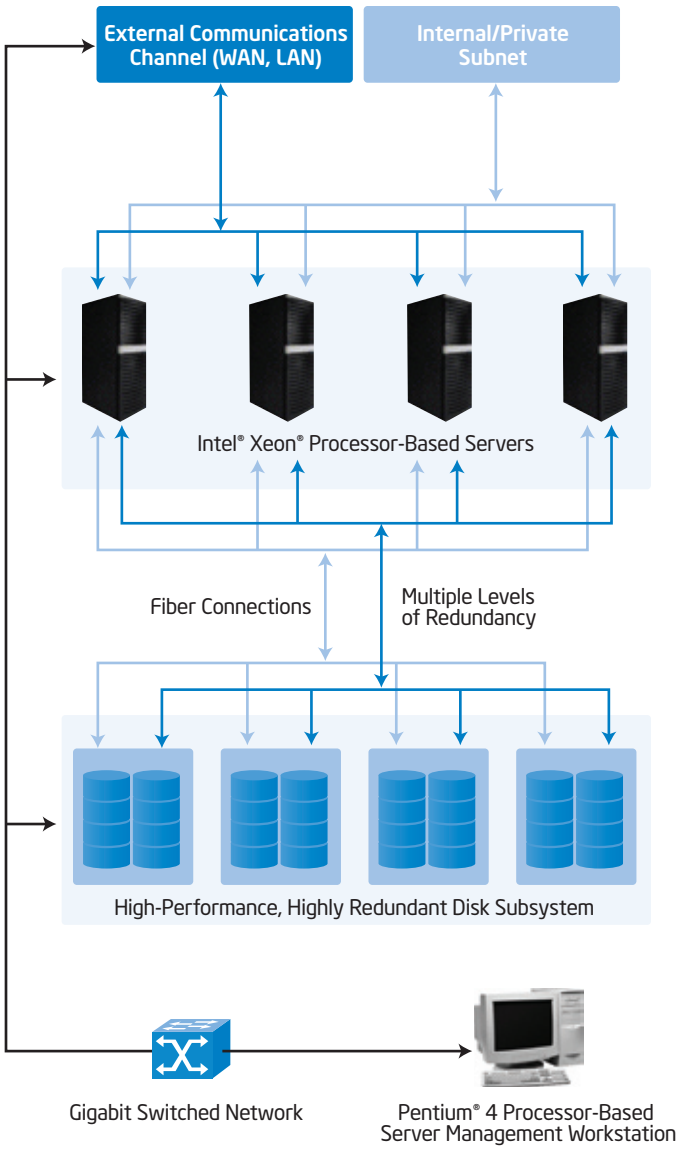
The system architecture is based on a shared-nothing approach and central processing unit (CPU) virtualization. Processors are dynamically allocated to applications based on business priority and a set of rules that adapt and optimize the database throughput to meet service level agreements.

The RDBMS software is based on Structured Query Language (SQL) that is American National Standards Institute (ANSI) compliant, and is capable of processing datasets in parallel. This capability enables the system to handle large data volumes and complex queries without degrading performance.

Security and Data Integrity

SOX required strict data classification and security management. Because executives were accountable for the data, they needed to know exactly what the numbers were and where they came from. Prior to our EDW effort, data silos in disparate datamarts created security and data integrity challenges. Our EDW effort helped meet these challenges.

Figure 1. Enterprise data warehouse solution architecture



The EDW consolidated data from several sources into a centralized repository for consistent management of data integrity. This gives Intel a single, trusted source for its information needs. Additionally, a highly integrated security model incorporated key controls to help protect data, yet was flexible enough to meet quickly growing business intelligence requirements for robust analytics. Figure 2 illustrates our security model.

Business Intelligence System Architecture

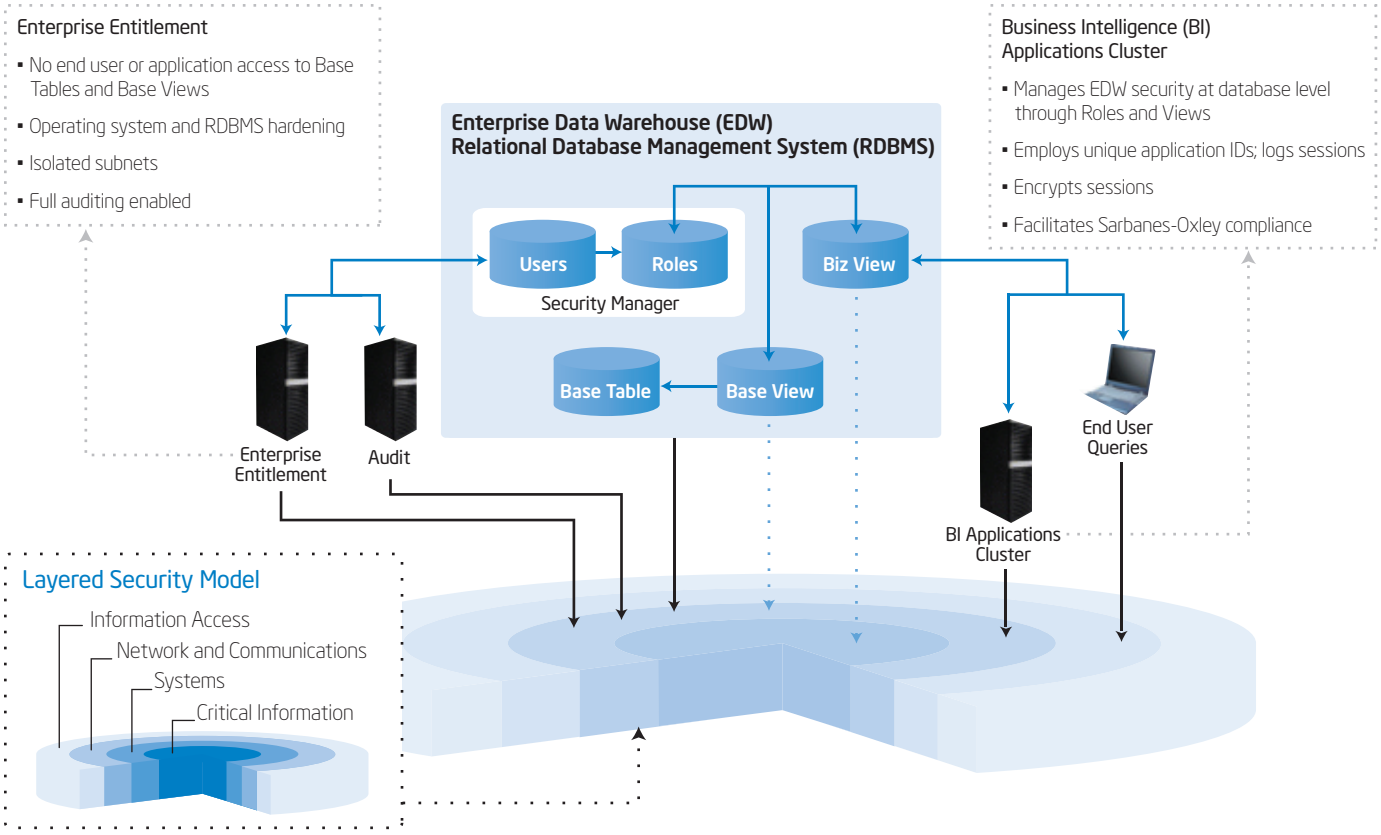
As part of the overall EDW architecture, we also specified a new BI system that could scale to meet users' needs as they found new ways to mine the data and discovered information that had essentially been hidden. The BI solution is comprised of Intel® Xeon® processor-based rack-mounted servers configured in a high-availability cluster with load balancing to take full advantage of system scalability and virtualization capabilities.

Deployment

By 2001, the team had defined the architecture and was ready to deploy the hardware. Deployment of the entire EDW was a staged process comprising hardware system installation, data model and metadata definition, data ET&L, BI system deployment, user training, and activation. We quickly deployed the hardware platform due to the plug-and-play nature of the components, a well-defined plan, and established vendor relationships. The more time-consuming part of the project involved defining the data model and metadata for the entire data warehouse, transforming the existing data into the new data model, loading it, and then building and deploying the applications.

We ran into several challenges during deployment, but they were not typical hardware or software integration problems. Most were process issues. We were integrating business-critical data and consolidating the processes of several business

Figure 2. Enterprise data warehouse security layers and capabilities



units under a single enterprise umbrella. We needed to create effective methods to manage the co-existence of applications on the system, so that when we brought in new applications with each step in the deployment, we didn't create conflicts with existing processes and applications.

One of the methods helps balance workloads across the system and all users at any given time. This prevents conflicts during peak usages when one application could potentially dominate resources, thus impacting the productivity of other users. We use a set of business rules to dynamically control the availability of resources to a specific user based on the workload impact on other users.

Results

This project was the most complicated BI implementation to date at Intel. It delivered substantial benefits and continues to allow Intel to reap significant rewards as the company leverages analytics capabilities to expose business value hidden in enterprise data.

Infrastructure Growth

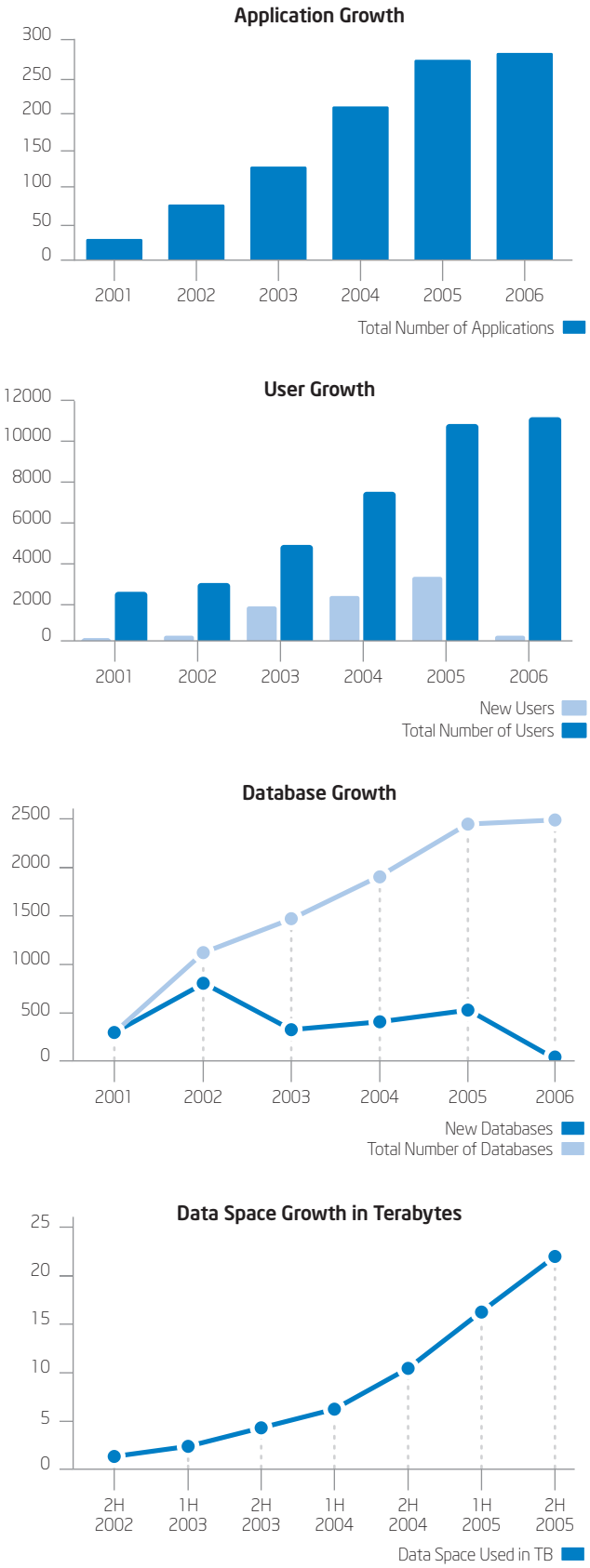
The system went online in 2001 with just over 100 users, 30 applications, and 317 databases occupying a few hundred gigabytes of data space. Over the next four years, Intel grew significantly. The EDW team added data and applications, put processes in place, and migrated additional business units to the new system (see Figure 3).

The team migrated business units following a standard governance and capacity model to make sure additions didn't create conflicts with existing processes and data sets, and to allow users to test the new system's output. By 2005, applications had grown 600 percent, users increased to over 11,000, databases increased six-fold, and our usable data space expanded to more than 20TB. Table 1, on page 11, summarizes the EDW at the end of 2005.

Today, Intel's EDW is a critical factor in the company's financial, customer, and supply chain processes. It provides new, detailed reporting capabilities, historical data for mining and analytics, and easily scales to meet growing business demands.

The EDW also helped Intel meet compliance requirements for finance and critical reporting applications under SOX. Intel divisions recognized the need for data quality and security, and the new system offers the technology needed to meet the stringent requirements of SOX. The EDW has become the record of reference (ROR) for business-critical and sensitive data, reflecting how trusted are the data and our processes to deliver accurate information for business needs or audits.

Figure 3. Enterprise data warehouse growth trends



Delivered Value

Early on, the EDW team decided to define a process to reflect the overall value of the program. Due to differences in IT costs versus direct business benefits or gains, we employed two models.

IT Cost Reduction and Avoidance

IT cost reduction and avoidance reflects financial impact by comparing the costs of managing individual and often redundant infrastructure versus the costs of landing new applications on the EDW. This value considers infrastructure management costs, benefits of data reuse, and other IT-related overhead costs. In 2005, IT calculated cost savings and cost avoidance of $22 million on IT infrastructure and overhead, as shown in Figure 4 on the next page.

IT cost savings and avoidance does not reflect the impact on direct business unit gains due to cost savings or company profitability derived from the benefits of deeper, multi-dimensional analytics—a capability that was severely limited before due to disparate databases and a lack of comprehensive BI infrastructure.

In-depth, Multi-dimensional Analytics

In-depth, multi-dimensional analytics and data mining provide visibility into financial activities at multiple levels and across many business units—all at once. End-to-end visibility from strong analytics enables managers to optimize their departments' functions in ways not previously possible with the old infrastructure. Analytics offered by the new system in managing price quotes, optimizing global procurement, and streamlining logistics provide good examples of the true business value derived from the highly integrated EDW/BI solution.

Price Quotes. Thousands of worldwide customers—from key customers to smaller buyers—send in tens of thousands of request for quotes (RFQs) for a product line of hundreds of products. The quote specialist must complete each quote in minutes, and do it effectively to ensure the company preserves its average selling price (ASP) and optimizes market segment share (MSS). Analytics are critical to this function to ensure profitability for the company.

By 2005, when we brought a particular quote department online with the EDW and BI applications, the quoting process was already well organized and effective. It considered over 20 aspects of the quote, and the specialist relied on multiple tools that accessed seven databases. But, the process did not necessarily translate into optimal pricing. Additionally, it was nearly impossible to produce consolidated monthly and quarterly reports that provided visibility across tens of thousands of quotes due to disparate databases and different data models.

This quoting process now benefits from the multi-dimensional analytics possible with BI applications and the EDW. The manager can clearly see how policies impact MSS, ASP, the supply chain, and profitability. The reports provide clear visibility across the entire product line in all regions of the world. They also give a clearer picture of how discounts are handled across large, key customers and smaller buyers so that the manager can fine tune the quoting process as often as necessary to optimize it.

Global procurement reporting. Global procurement reporting (GPR) at Intel derived a similar benefit. Prior to implementing our new EDW with BI applications, GPR used multiple tools across several systems that provided a one-dimensional view into the supply chain. Enterprise resource planning (ERP) depended on other staff and analysts in the company to collect, transform, and analyze the data for reports on their activities. With the new system, all material professionals have access to the same data. They can perform in-depth, multi-view analyses across commodities, geographies, suppliers, and more. We incorporated the EDW and BI system into a new e-Procurement process that uses information derived from the EDW.

Materials movement. The process of moving products and materials around the world also benefited from better analysis across multiple geographies and shippers. In 2004, the EDW enabled logistics managers to more effectively analyze shipping and handling costs, reorganize schedules, and renegotiate contracts with shippers to reduce operating costs.

These are just a few examples of how Intel has benefited from a new EDW/BI solution. These examples and others relate to the overall impact the EDW has had on Intel divisions resulting in hundreds of millions of dollars in business value.

Conclusion

Our new Intel® architecture-enabled EDW and BI system gives our business units important new analytic capabilities that scale to meet demand using a single, trusted source for their data needs. This resulted in bottom-line benefits in the hundreds of millions of dollars from 2001 through 2005. It also delivered IT infrastructure cost avoidance in the tens of millions of dollars by consolidating databases and integrating data into a unified RDBMS.

The new data security and BI reporting capabilities enable us to provide a highly secure enterprise database that delivers excellent performance to meet critical service level agreements and help Intel with SOX regulatory reporting. The EDW solution continues to easily scale to provide Intel's business units with the timely, accurate information they need to grow and optimize their business.

Authors

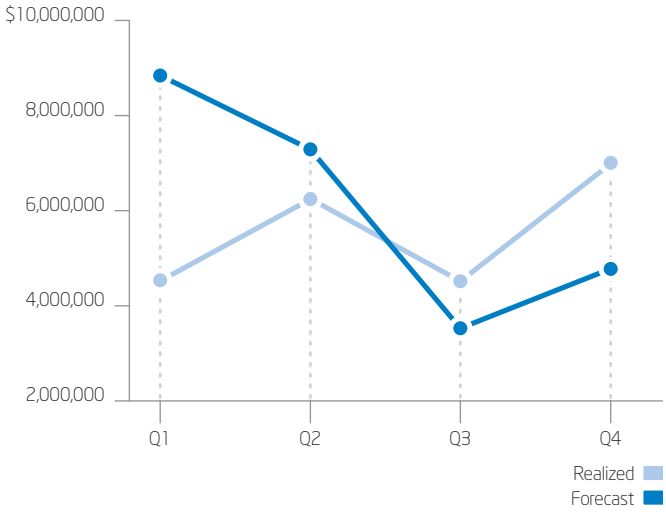
Charles Eden is a systems engineering manager with Intel IT.

Venky Padmanabhan is a capability architecture manager with Intel IT.

Table 1. 2005 enterprise data warehouse statistics

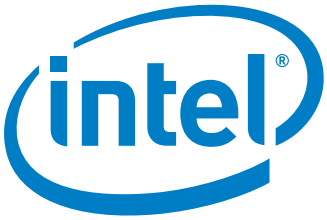
Item	2005
Usable database size	23TB
Average number of SQL statements per day	1.3 million+
Database users	11,000+
Business intelligence tool users	20,000+
Average number of reports per day	10,000+
Total number of rows loaded per day	116,557,463+
Total number of bytes loaded per day	33,286,084,480
Total number of reports delivered annually	5 million+
Production uptime/system availability	99.99%

Figure 4. Cost savings and avoidance in 2005



Acronyms

ANSI	American National Standards Institute
ASP	average selling price
BI	business intelligence
CPU	central processing unit
EDW	enterprise data warehouse
ERP	enterprise resource planning
ET&L	extract, transform and load
GPR	global procurement reporting
MPP	massively parallel processing
MSS	market segment share
RDBMS	relational database management system
RFQ	request for quote
ROR	record of reference
SQL	Structured Query Language
SOX	Sarbanes-Oxley Act of 2002



www.intel.com/IT

Printed in USA ♻ Please Recycle

0206/SHW/RDA/XX/PDF Order Number: 311290-001US